

# VI Jornadas del SID@R

The logo for SID@R, featuring a stylized blue figure in a wheelchair with a green '@' symbol on the wheel, followed by the text 'SID@R' in blue and green.

**RDF/XML y la tendencia de los  
metadatos orientados al  
contenido:**

**Implicaciones para la  
accesibilidad**

**Eva M<sup>a</sup> Méndez Rodríguez**  
emendez@bib.uc3m.es

**Universidad Carlos III  
de Madrid**



**Dpto. Biblioteconomía y  
Documentación**

# Contenidos

- ◆ Problemas en el acceso a la heterogeneidad de contenidos Web ¿SW?.
- ◆ Metadatos: concepto, uso y tipología
- ◆ Tipos de metadatos:
  - ◆ Orientados al recurso.
  - ◆ Orientados al contenido: NKOS
- ◆ ¿Son los metadatos una moda o una necesidad? Metadatos y accesibilidad.
- ◆ La importancia del vocabulario en el acceso a la información.
- ◆ *Schemas* de metadatos y *schemes* de contenido.
- ◆ NKOS y Web semántica: el acceso a la información por conceptos: Thesaurus vs. Ontologías.
- ◆ RDF / XML y la creación de NKOS para acceder al contenido.

## **Problemas en el acceso a la heterogeneidad de contenidos Web**

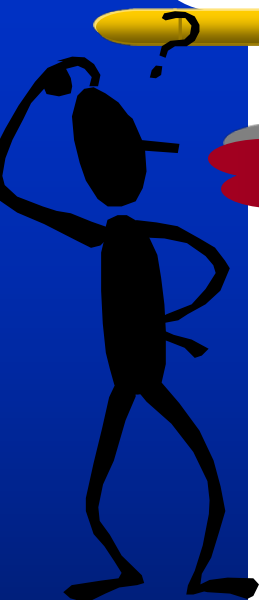
---

# Implicaciones de la WWW



**V**olumen **V**ariiedad **V**olatilidad

# La Web Semántica está despegando



ESTÁNDARES

Z39.50

PROTOCOLOS

xml  
XSL

RDF  
Metadata

《EAD》

METADATOS

  
Dublin Core

SW

PERFILES DE  
APLICACIÓN



# Heterogeneidad de información en la Web

Recursos de  
información  
en Red

Bases de datos

Bibliotecas  
digitales

Imágenes

Colecciones de  
museos

E-Publicaciones

Libros tradicionales

OTROS



# Problemas

- ◆ Grandes conjuntos de datos
  - ✦ Los formatos son heterogéneos y complejos.
  - ✦ El coste del almacenamiento.
  - ✦ El ancho de banda se amplía pero no directamente proporcional a la velocidad de almacenamiento y recuperación de los datos.
- ◆ Las colecciones de datos y documentos digitales (DLO) son:
  - ✦ Grandes y diversas.
  - ✦ Numerosas.
  - ✦ Cambian rápidamente.

# Problemas para el acceso al contenido

## ◆ Múltiples...

P  
R  
O  
B  
L  
E  
M  
A

- ◆ Entornos informativos / disciplinas.
- ◆ Modelos de datos y soluciones *ad hoc*.
- ◆ Idiomas y conjuntos de caracteres diversos.
- ◆ Esquemas de metadatos (DC, GILS, FGDC, LOM, IMS, etc).
- ◆ Protocolos de búsqueda (HTTP, SOAP, OIA, Z39.50, etc.).
- ◆ Formatos de fichero (html, xml, pdf, jpg, png, etc.).
- ◆ Vocabularios. (LCSH, MESH, AAT, etc.)

**Accesibilidad**

- Elaborar sintaxis de interrogación.
- Crear herramientas de alto nivel para la descripción semántica.
- Se necesitan Sistemas de Organización del conocimiento (ontologías) .



# Metadatos: concepto, uso y tipología

---

## Metadatos son...

### ... **Datos**

... Datos sobre datos

... **Descripciones estructuradas** de un objeto de información.

... Un **conjunto de elementos** y atributos para caracterizar la información.

... **Estándares** semánticos para describir objetos de datos discretos, que precisan de una semántica y estructura.

### Su finalidad es...

... Describir y recuperar información Web.

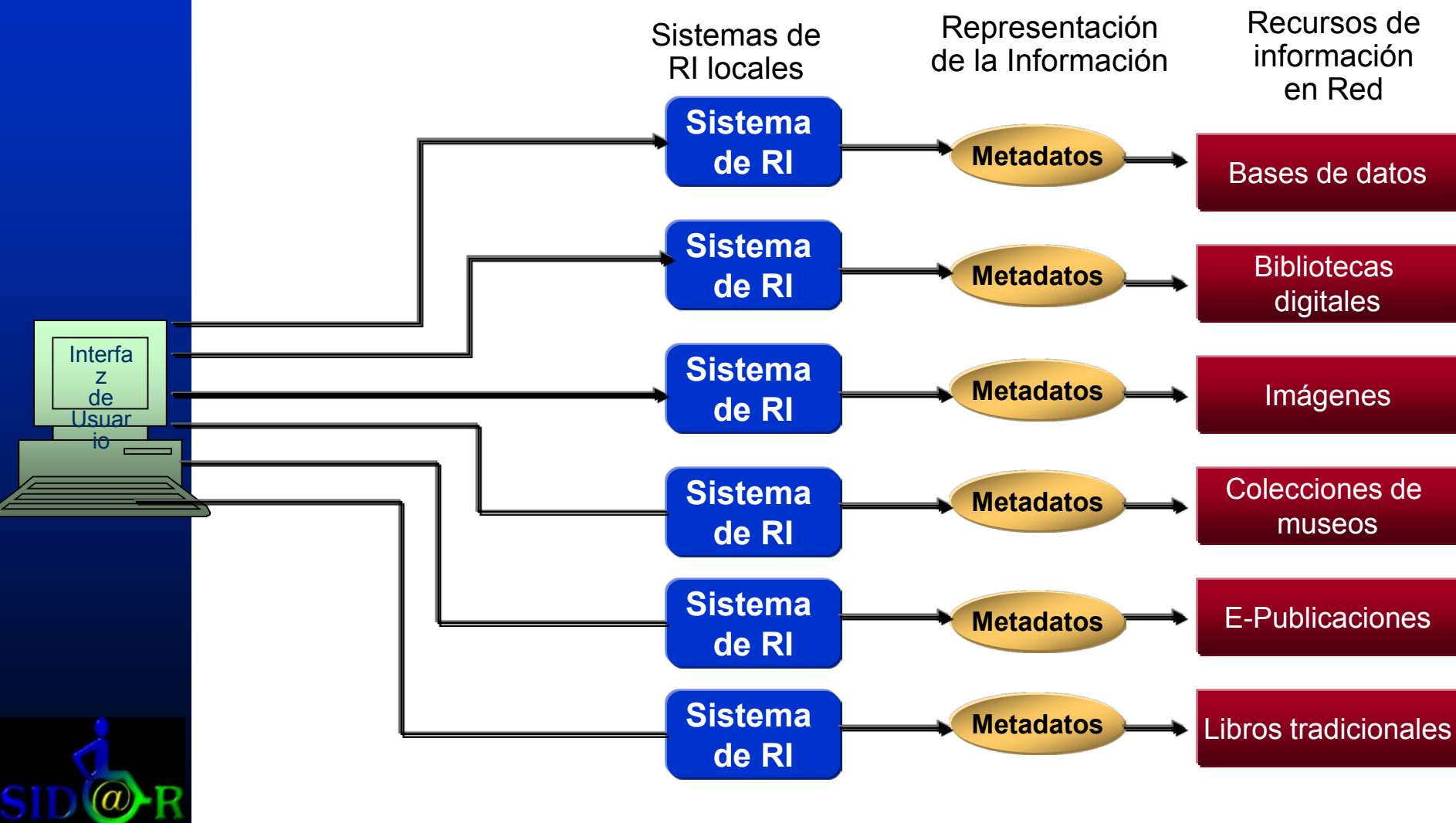
... Fundamentar SRI de sistemas y servicios de información digital.

## Contexto general de los metadatos

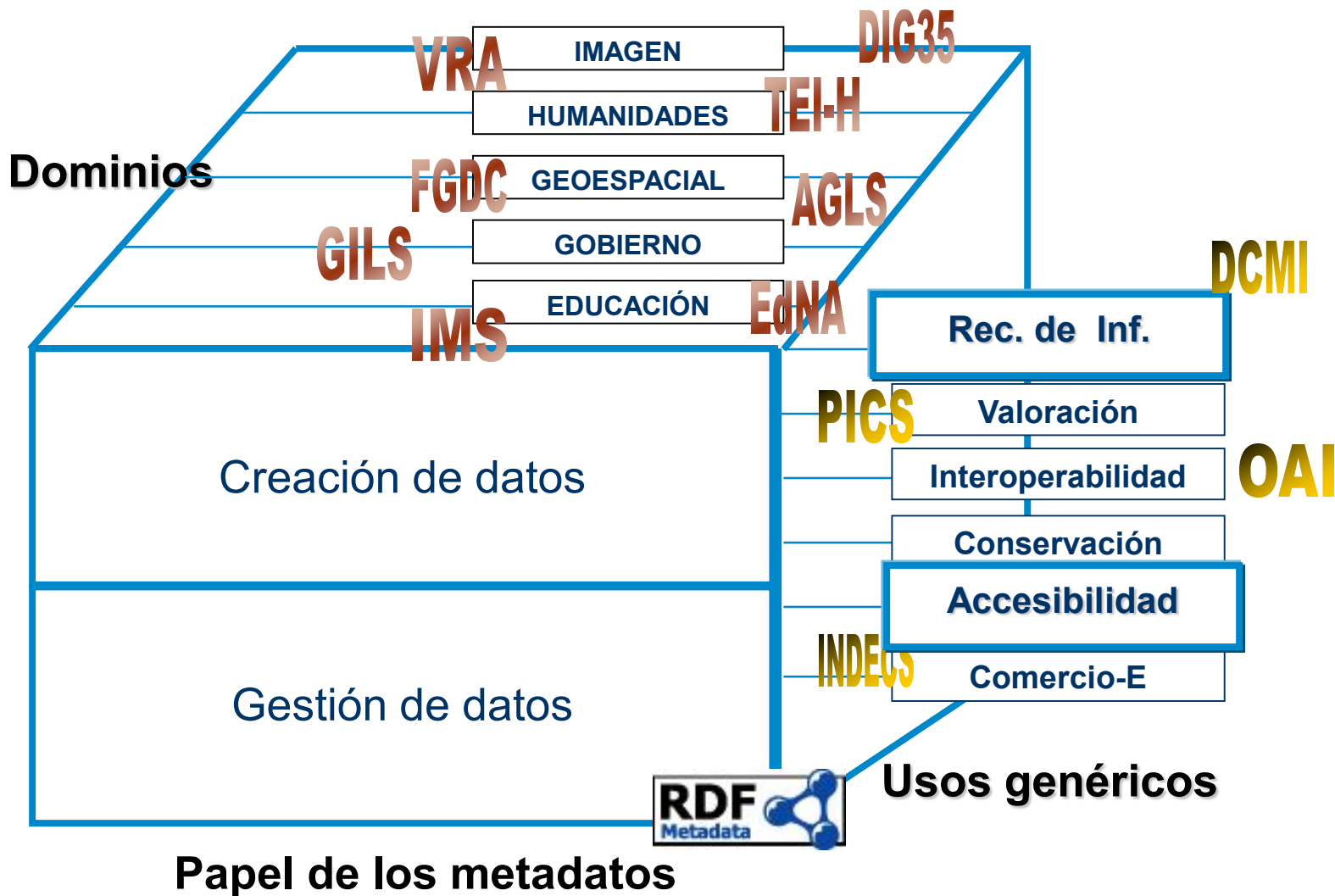


**La estructura y semántica de los modelos de metadatos aplicados a la recuperación es factible, aplicable y operativa en entornos de información abarcables, definidos y finitos.**

# RI con Metadatos



# Metadatos: aplicación y tipología



# Tipos de metadatos

## ◆ Dependientes del recurso

**Catalogación**

- ◆ Modelos semánticos de representación de las características básicas de un DLO

**Indización**

## ◆ Dependientes del contenido

- ◆ Modelos de representación de las materias y de las relaciones entre ellas (Tesauros, ontologías... )

# Metadatos orientados al contenido

- ◆ Metadatos basados en el contenido directo:
  - ✦ (ej. Índices de contenido, color de imágenes).
  
- ◆ Metadatos descriptivos del contenido:
  - ✦ Independientes de un dominio específico (p.ej. Título, autor...)
  - ✦ Dependientes del dominio (p. ej. Cobertura geográfica en FGDC)

# Metadatos ¿moda o necesidad?

- ◆ Descripción de recursos
  - Representación estructurada de DLOs.
  - Elementos representan y describen atributos de un DLO.
- ◆ Implicaciones para la accesibilidad
  - Punto de verificación 13.2 (WCAG)
  - Utilización de metaetiquetas.
  - **Seleccionar de forma consistente el contenido de los metadatos:** NKOS (*Network Knowledge Organization System/Services*).



# **La importancia del vocabulario en el acceso a la información**

---

# La necesidad del control de vocabulario



# Sin control terminológico.....

## ◆ Los usuarios de información

- ◆ Utilizan términos de búsqueda incorrectos.
- ◆ No consiguen encontrar significados relevantes.
- ◆ Sufren la inflación de la información
- ◆ Los mejores resultados serían semejantes a un buen uso de AltaVista.
- ◆ El acceso a contenidos relevantes estará en función del idioma y de la “suerte” (*serendipity*)

## ◆ Los creadores de información

- ◆ Describen recursos de manera inconsistente
- ◆ No pueden especificar conceptos jerárquicos
  - España es una país, España está en Europa, Europa está/es...
- ◆ No pueden perpetuar el crecimiento lingüístico
- ◆ Se produce un proceso de creación de información aislado.

## Con control terminológico...

### ◆ Los usuarios pueden:

- ◆ Mejorar el acceso efectivo a los recursos electrónicos.
- ◆ Mejorar la eficacia en el acceso a los recursos.
- ◆ Reducir el número de resultados de búsqueda irrelevantes.
- ◆ Encontrar lo que están buscando.
- ◆ Incluso aprenden a pensar y ejecutar sentencias de búsqueda de una forma más estructurada.

### ◆ Los creadores pueden:

- ◆ Producir recursos de mayor calidad semántica.
- ◆ Introducir conceptos semántica y estructuralmente más complejos.
- ◆ Involucrarse en procesos de integración de terminologías a nivel nacional, internacional o global.
- ◆ Integrar eficazmente tanto los nuevos recursos/conceptos como los preexistentes en el sistema de información.

# ***Schemas de metadatos y schemes de contenido***

---

# Schema= esquema de estructura

- *Schema* es un documento que describe un vocabulario XML o RDF
- En sentido más genérico un *schema* ("de metadatos") es el conjunto de elementos y las reglas que constituyen un modelo de metadatos.
- Los esquemas de metadatos son modelos, formatos, estándares, etc. para estructurar la información sobre la información
- Un ejemplo: DCMI (15 elementos básicos para la descripción de RI).

# Metabrowser (schema DC)

**Generic Dublin Core Metabrowser Tutorial**

File Edit View Metadata Resources Macros Help

Metadata Schemas  
Metadata Controlled Lists (Schemes)  
Metadata Templates

Default Domain: Dublin Core

Dublin Core

- DC.Contribu
- DC.Coverag
- DC.Creator
- DC.Date
- DC.Description
  - DC.Description
  - DC.Description.abstract
  - DC.Description.tableOf
- DC.Format
- DC.Identifier
- DC.Language
- DC.Publisher
- DC.Relation
- DC.Rights
- DC.Source
- DC.Subject
- DC.Title
- DC.Type

Home

Contact us

Generic Dublin Core Metabrowser Tutorial

This Tutorial uses generic Dublin Core elements as perscribed by [Metadata Initiative](#)

Edit is Active

Name: DC.Description

Content:

Scheme:

Name	Content	Scheme
DC.Identifier	http://metabrowser.spirit.net.au/tutorial/DC.html	URI
<b>DC.Description</b>		
DC.Title	Generic Dublin Core Metabrowser Tutorial	
DC.Subject	Dublin Core; Metadata; Creation Tools; Metabrowser	
DC.Description	Tutorial for learning Metadata Creation using Metabrowser	
DC.Creator	Spirit Consulting	
DC.Date	2001	ISO8601

Current File: <No File> Modified 26/10/01

## Scheme=esquema de contenido

- ◆ Un *scheme* es un fichero de vocabulario controlado, la lista de valores posibles que puede tener una metaetiqueta en concreto.
- ◆ Conjunto normalizado de valores que pueden tener los schemas.
- ◆ Un ejemplo (o dos):
  - MeSH (materias en medicina <DC.Subject>
  - ISO 8601 (estándar de fechas <DC.Date>



# Metabrowser (scheme EdNa)

The screenshot shows the Metabrowser application window titled 'Universidad Carlos III de Madrid'. The 'Metadata Resources' menu is open, showing 'Metadata Schemas', 'Metadata Controlled Lists (Schemes)', and 'Metadata Templates'. The 'edna.directory' scheme is selected in the 'Metadata Controlled Lists (Schemes)' menu. The 'Default Domain: Dublin Core' is also visible.

The left sidebar shows a tree view of metadata elements. The 'edna.directory' scheme is selected, and its contents are displayed in the main area. The 'edna.directory' scheme is highlighted in the 'Metadata Controlled Lists (Schemes)' menu.

The main area displays the 'edna.directory' scheme. The 'Name' field is 'EDNA.CategoryCode', the 'Content' field is empty, and the 'Scheme' field is 'edna.directory'.

The bottom table shows the metadata elements for the 'edna.directory' scheme:

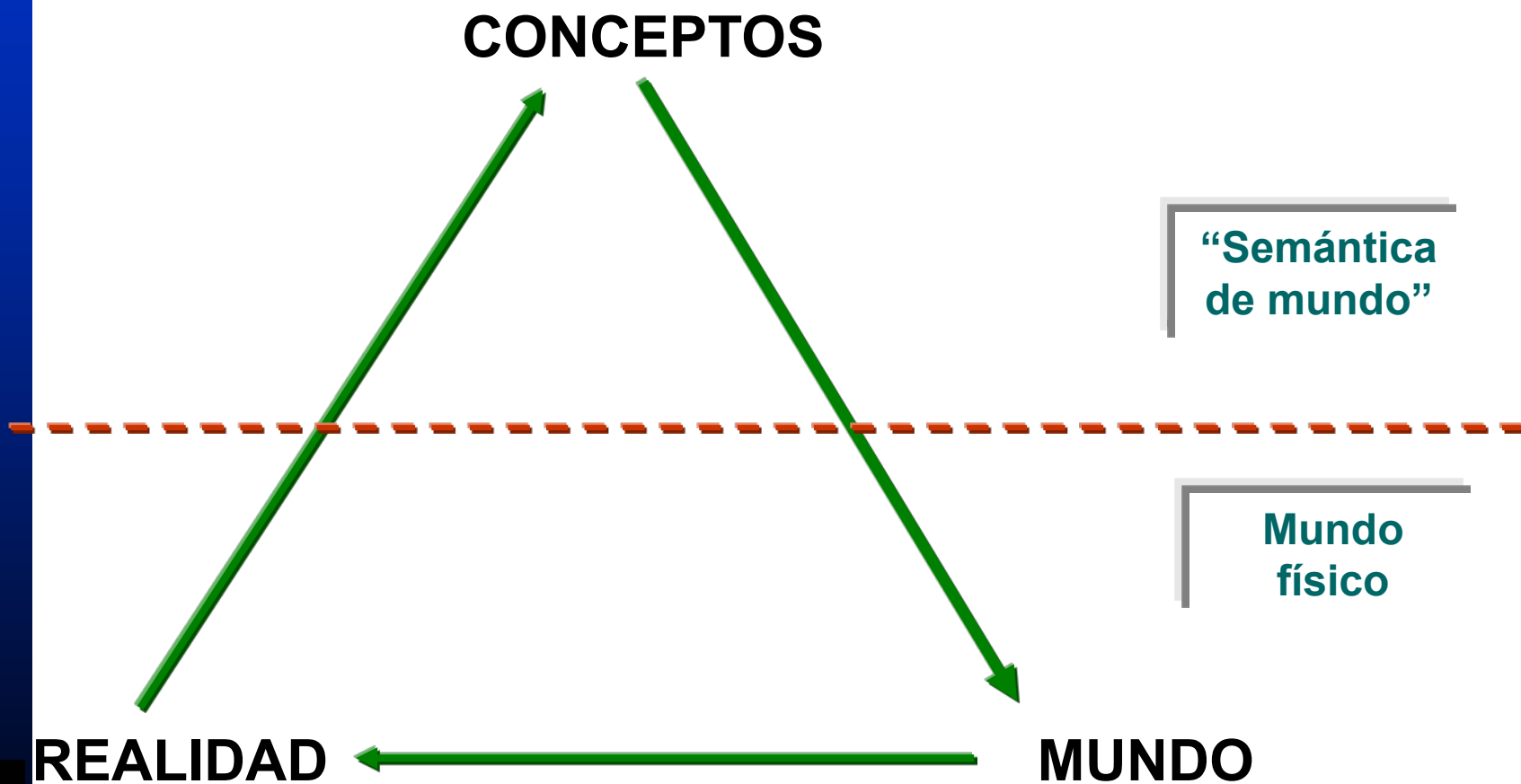
Name	Content	Scheme
Content-Type	text/html; charset=iso-8859-1	
DC.Description		
EDNA.Categ...		edna.directory
description	La Universidad Carlos III de Madrid ofrece un modelo de educación integral, ...	
Keywords	universidad, university, educacion, education, madrid, administracion y direcc...	

The status bar at the bottom shows 'Current File: <No File>' and 'Modified'.

# **NKOS y Web semántica: el acceso a la información por conceptos. Thesaurus vs. Ontologías.**

---

# Dimensiones del conocimiento



## SW: una visión desde la Documentación

- ◆ No es la W2: es la Web.
- ◆ Implica crear un interfaz común a todos los conceptos y bases de conocimiento terminológico en la Web.
- ◆ Sistema de información de acceso integrado de datos, términos y conceptos
- ◆ Aunar distintos recursos que existen en diferentes sistemas o servicios de información digital basados en Web (diccionarios, tesauros, esquemas de clasificación, etc).

## SW implica...

- ◆ Espacios de nombre (*xmlns*) para identificar un *schem[a/e]* y cada elemento dentro de su ***schema*** y cada concepto dentro de cada ***scheme***.
- ◆ Un amplio esquema conceptual que pueda mantener todas los tipos de información semántica y especificar las relaciones entre ellos (*ontologías*).
- ◆ Definición de servicios para interactuar con el sistema.

## Conclusiones / Reflexiones

---

# Cómo garantizar el acceso al contenido...

## Factores clave para la creación de contenidos accesibles en la SW



**GLOBALIZACIÓN**

**INTEROPERABILIDAD**

**NORMALIZACIÓN**

**COMPARTICIÓN DE DATOS**

## Conclusiones

- ◆ La consistencia en el acceso a la información por conceptos y su representación en RDF mejorará la accesibilidad de la información en dos sentidos:
  - ◆ La capacidad de los metadatos de anticiparse al contenido/condiciones de acceso de un DLO.
  - ◆ La posibilidad de crear un perfil específico de aplicación de los metadatos a la accesibilidad.
- ◆ Necesaria investigación/inversión en metadatos y en organización de conocimiento.
- ◆ Principales problemas/objetivos:
  - ◆ **ACCESIBILIDAD**
  - ◆ **INTEROPERABILIDAD**
- ◆ RDF es a la WS, lo que HTML a la Web.



## Conclusiones

**EL FUTURO DE LA WEB  
SEMÁNTICA ESTÁ  
POR LLEGAR...**



## Discusión, preguntas...



Tenerife, 24 de octubre 2002